

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/50457/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Rostami-Tabar, Bahman ORCID: <https://orcid.org/0000-0002-3730-0045>, Babai, M. Zied, Syntetos, Argyrios ORCID: <https://orcid.org/0000-0003-4639-0756> and Ducq, Yves 2013. Demand forecasting by temporal aggregation. Naval Research Logistics 60 (6) , pp. 479-498. 10.1002/nav.21546 file

Publishers page: <http://dx.doi.org/10.1002/nav.21546>
<<http://dx.doi.org/10.1002/nav.21546>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Demand Forecasting by Temporal Aggregation

Abstract: Demand forecasting performance is subject to the uncertainty underlying the time series an organisation is dealing with. There are many approaches that may be used to reduce uncertainty and thus to improve forecasting performance. One intuitively appealing such approach is to aggregate demand in lower-frequency ‘time buckets’. The approach under concern is termed to as Temporal Aggregation and in this paper we investigate its impact on forecasting performance. We assume that the non-aggregated demand follows either a moving average process of order one or a first-order autoregressive process and a Single Exponential Smoothing (SES) procedure is used to forecast demand. These demand processes are often encountered in practice and SES is one of the standard estimators used in industry. Theoretical Mean Squared Error expressions are derived for the aggregated and non-aggregated demand in order to contrast the relevant forecasting performances. The theoretical analysis is supported by an extensive numerical investigation and experimentation with an empirical dataset. The results indicate that performance improvements achieved through the aggregation approach are a function of the aggregation level, the smoothing constant and the process parameters. Valuable insights are offered to practitioners and the paper closes with an agenda for further research in this area.

Keywords: Demand Forecasting, Temporal Aggregation, Stationary Processes, Single Exponential Smoothing

1. INTRODUCTION

Demand uncertainty is among the most important challenges facing modern companies [1]. The existence of high variability in demand for fast moving and slow/intermittent moving items (items with a high ratio of zero observations) pose considerable difficulties in terms of forecasting and stock control. Deviations from the degree of variability accommodated by the Normal distribution often render standard forecasting and inventory theory inappropriate [2, 3, 4].

There are many approaches that may be used to reduce demand uncertainty and thus to improve the forecasting (and inventory control) performance of a company. An intuitively appealing such approach that is known to be effective is demand aggregation [5]. One possibility is to aggregate demand in lower-frequency ‘time buckets’ (for example aggregate weekly data into monthly) and such a time series transformation approach is often referred to, in the academic literature, as *Temporal Aggregation* [6]. Another aggregation approach discussed in the literature and often applied in practice is the *Cross-Sectional* or *Contemporaneous Aggregation*, which involves aggregating different time series in order to improve performance across a group of items [7]. Such an approach is equivalent to aggregating data for one single Stock Keeping Unit (SKU) across a number of depots or stock locations. Natural, practically useful, associated forms of aggregation involve also geographical consolidation of data or aggregation across markets.

Although no empirical studies exist that document the extent to which aggregation takes place in practical settings, this is an approach that is known to be popular amongst practitioners not the least because of its intuitive appeal. In practical terms, the benefit will depend on the type of aggregation and of course the data characteristics. Cross-sectional aggregation for example usually leads to variance reduction. This is due to the fact that fluctuations in the data from one time series may be offset by the fluctuations present in another time series [8]. Contrary to cross-sectional aggregation, in temporal aggregation, variance is increased. However, it is easy to show that temporal aggregation can reduce the coefficient of variation of demand. In any case, the implied benefit coupled with the ease of implementing such approaches renders them a popular choice in industry.

From an academic perspective the emphasis to date has been mainly on cross-sectional aggregation. Moreover, and although most inventory forecasting software packages support aggregation of data, this would also typically cover cross-sectional aggregation only. The consideration of temporal aggregation has been somewhat neglected by software manufacturers and academics alike despite the potential opportunity for adding more value to real world practices. In this paper we aim to advance the current state of knowledge in the area of demand forecasting temporal aggregation.

Demand data may be broadly categorized as intermittent and fast. Aggregation of demand in lower-frequency ‘time buckets’ enables the reduction of the presence of zero observations in the former case or, generally, reduce uncertainty in the latter. Intermittent demand items (such as spare parts) are known to cause considerable difficulties in terms of forecasting and inventory modeling. The presence of zeroes has significant implications in terms of: i) difficulty in capturing underlying time series characteristics and fitting standard forecasting models; ii) difficulty in fitting standard statistical distributions such as the Normal; iii) deviations from standard inventory modeling assumptions and formulations – that collectively render the management of these items a very difficult exercise. Temporal aggregation is known to be applied widely in military settings (very sparse data), the after sales industry (service parts) etc. Recent empirical studies in this area [6, 9] have resulted in some very promising results pointing out also the need for more theoretical analysis. Although the area of forecasting with temporal aggregation in an intermittent demand context is a very interesting one both from an academic and practitioner perspective, in this paper we consider only the most often occurring cases of fast demand items. Analysis in an intermittent demand context is an important avenue for further research and this issue is discussed in more detail in the last section of the paper.

1.1. Objectives and Organization of the Paper

In this paper we study analytically the effects of temporal aggregation on forecasting when the underlying series follow an Autoregressive process of order one, AR(1) or a Moving Average process of order one, MA(1) and the forecasting method is the Single Exponential Smoothing (SES). Both assumptions bear a significant degree of realism. As it will be discussed later in the paper there is evidence to support the fact that demand often follows the stationary processes assumed in this work (43% of the empirical series available in our research follow such processes). Moreover, SES is a very popular forecasting method in industry [10, 11, 12, 13]. Although its application implies a non-stationary behavior of the demand, sufficiently low smoothing constant values introduce minor deviations from the stationarity assumption whilst the method is also unbiased.

In this work we compare the variance of the forecast error (or equivalently, by considering an unbiased estimation procedure, the mean square error) obtained based on the aggregated demand to that of the non-aggregated demand. Comparisons are performed at the original (non-aggregate) demand level. We mathematically show that the ratio of the Mean Squared Error (MSE) of the latter approach to that of the former is a function of the aggregation level, the process parameters and the exponential smoothing constant. The mathematical analysis is complemented by a numerical investigation to evaluate in detail the conditions under which aggregation leads to forecast performance improvements. Next, we validate empirically our theoretical results (by means of simulation on a dataset provided by a European superstore) and by doing so we also offer some very much needed empirical evidence in the area of temporal aggregation. Finally, important managerial insights are derived and tangible suggestions are offered to practitioners dealing with inventory forecasting problems.

To the best of our knowledge, the only papers directly relevant to our work are those by Amemiya and Wu [14] and Tiao [15] for the AR and MA process respectively. In both cases the researchers investigated the forecast performance of temporal aggregation strategies under an (Auto-Regressive Integrated Moving Average) ARIMA-type framework. However, the results presented in these two papers remain preliminary in nature while the experimental setting may also be criticized in terms of the estimation procedures considered [16]. In addition, no empirical results were provided. Important as they are, both papers focused on characterizing the aggregated demand series rather than the forecast performance. These issues are further discussed in the next section of the paper.

This paper attempts to fill this gap and provides helpful guidelines to select the appropriate approach under such demand processes. The work discussed in this paper can be extended to analyze more general cases such as $AR(p)$, $MA(q)$ or indeed $ARMA(p,q)$ processes. However, the analysis and presentation of results would become complex. Since our main objective is to obtain some key managerial insights, we shall restrict our attention to the AR(1) and MA(1) processes only.

The remainder of this paper is structured as follows. In Section 2 we provide a review of the literature on the issue of temporal demand aggregation. In Section 3 we describe the assumptions and notations used in this study. In Section 4, we conduct an analytical evaluation of the MSE related to both the aggregation and non-aggregation approaches. The conditions that determine the comparative performance of the two approaches are determined in Section 5 followed by an empirical analysis conducted in Section 6. The paper concludes in Section 7 with the implications of our work for real world practices along with an agenda for further research in this area.

2. LITERATURE REVIEW

In the supply chain and demand planning literature, demand aggregation is generally known as a 'risk-pooling' approach to reduce demand fluctuation for more effective material/capacity planning [1]. Demand uncertainty may considerably affect forecasting performance with further detrimental effects in production planning and inventory control. It has been shown by Theil [17], Yehuda and Zvi [18], Aigner and Goldfeld [19] that demand uncertainty can be effectively reduced through appropriate demand aggregation and forecasting.

Temporal aggregation in particular refers to the process by which a low frequency time series (e.g. quarterly) is derived from a high frequency time series (e.g. monthly) [6]. This is achieved through the summation (bucketing) of every m periods of the high frequency data, where m is the aggregation level. There are two different types of temporal aggregation: non-overlapping and overlapping. In the former case the time series are divided into consecutive non-overlapping buckets of time. In the overlapping case, at each period the oldest observation is dropped and the newest is included. In both the non-overlapping and overlapping cases, the length of the time bucket equals the aggregation level. In this paper only the case of temporal non-overlapping aggregation is considered; overlapping aggregation is an issue left for further research and this is discussed in more detail in the last section of the paper.

Most of the literature that deals with temporal aggregation may be found in the Economics discipline. The analysis of temporal aggregation starts with the work of Amemiya and Wu [14]. They have shown that if the original variable follows a p^{th} order autoregressive process, $AR(p)$, then the non-overlapping aggregates follow a mixed autoregressive moving average (ARMA) model of the (p,q) form where $q^* = [(m-1)(p+1)/m]$. By considering the ratio of MSE of disaggregate and aggregate prediction (3 linear predictors were considered) at the aggregated level, they have shown that the MSE of non-aggregated forecasts is greater than that of the aggregated ones, i.e. the aggregation approach outperforms the non-aggregation one. Tiao [15] has investigated the effect of non-overlapping temporal aggregation on a non-stationary process of the Integrated Moving Average $IMA(d,q)$ form, where d is the integrated parameter and q is the moving average parameter. He has shown that the aggregated process is of the $IMA(d,q^*)$ where q^* is $q^* \leq [d+1+(q-d-1/m)]$. They applied a conditional expectation to obtain one step ahead forecasts at the aggregated level based on the non-aggregated and aggregated series. Subsequently, the efficiency of the aggregated forecasts was defined as the ratio of the variance of the forecast error of the non-aggregated to the aggregated series when the aggregation level is large. They have shown that when $d=0$ the ratio under concern equals to 1 and the comparative benefit of using the non-aggregated forecasts is increasing with d .

Our work considers the case of $AR(1)$ and $MA(1)$ processes and as such some of the theoretical results presented in the above discussed papers are of direct relevance to our analysis. Our work differs from these papers though and extends them in some very significant ways: i) optimal estimators are seldom used in practice not only due to the computational requirements that are typically prohibitive but also the lack of understanding on the part of the managers of their functionality. In addition, there is evidence to support the fact that simple forecasting methods (such as SES that is used in our work) perform at least as good as more complex theoretically coherent alternatives [20]; ii) a difficulty associated with aggregation methods is the fact that a disaggregation mechanism is also required since very often forecasts are needed at the original/non-aggregate demand level. Both papers consider a comparison at the aggregate level which addresses only part of the forecasting problem. Consideration of a comparison at the original demand level, which is the case considered in our work, addresses

another part of the problem and is an important extension of the research already being done¹; iii) no empirical analysis has been undertaken in both papers in contrast with our work where our findings are empirically validated; iv) we complement our analysis by means of further numerical investigations to identify the optimum aggregation level and smoothing constant values that need to be used.

Brewer [21] studied the effects of non-overlapping temporal aggregation on ARMA (p, q) processes. He showed that aggregating such processes results in ARMA processes with autoregressive order p and moving average order r , ARMA (p, r), where $r = [p + 1 + (q - p - 1/m)]$. Stram and Wei [22] have studied the relationship between the autocovariance function of non-aggregated and aggregated processes. They have shown that the autocovariance function of the latter can be computed based on the autocovariance function of former; in particular the autocovariance function after aggregation is a function of the aggregation level and autocovariance function before aggregation. Souza and Smith [23] showed that for AR Fractionally IMA (ARFIMA) models temporal aggregation results in bias reduction.

Athanasopoulos et al. [24] have recently looked at the effects of non-overlapping temporal aggregation on forecasting accuracy in the tourism industry. They have conducted an empirical investigation using 366 monthly series and some forecasting methods tested in the M3 competition data [20], namely Innovations state space models for exponential smoothing (labeled ETS), the ARIMA methodology, a commercial software (Forecast Pro), damped trend [25], the Theta method and naïve. The monthly series were aggregated to be quarterly, and the quarterly series were further aggregated to be yearly. Subsequently, they compared the accuracy of the forecasts made before and after aggregation. They considered one and two step-ahead forecasts and three statistical measures were used to compare the results: Mean Absolute Percentage Error (*MAPE*), Mean Absolute Scaled Error (*MASE*) and Median Absolute Scaled Error (*MdASE*). The aggregated forecasts at the yearly level (whether produced from monthly or quarterly data) were found to be more accurate than the forecasts produced from the yearly data directly. This study provided considerable empirical evidence in support of temporal aggregation.

Luna and Ballini [26] have used a non-overlapping aggregation approach to predict daily time series of cash money withdrawals in the neural forecasting competition, NN5². Each time series consisted of 735 daily observations which have been used to forecast 56 daily steps ahead for two sets of 11 and 111 time series. Daily samples were aggregated to give weekly time series and then an adaptive fuzzy rule-based system was applied to provide 8-step-ahead forecasts (thus aggregation reduced the forecast horizon from 56 to 8 steps). Two different aggregation approaches were evaluated for this purpose: the historical top-down (TD-H) approach and the daily top-down (TD-DM) approach, where the main difference between the two was the disaggregation procedure. In the former case aggregated forecasts were dis-aggregated based on historical percentages. In the latter case, the daily estimations were ‘corrected’ by multiplying them by the associated weekly estimation and dividing by the sum of the seven daily estimated samples. The symmetric MAPE (*sMAPE*) and the Mean Absolute Error (*MAE*) were used to compare the results. The researchers showed that the aggregated forecasts produced by the two approaches performed similarly or better than those given by the daily models directly. The reduction of a forecast horizon from 56 to 8 steps ahead would be intuitively expected to lead to performance improvement.

Although many papers consider the case of fast moving items or continuous-valued time series, integer time series have received less attention in a temporal aggregation context. Brannas et al. [27] first studied the non-overlapping temporal aggregation of an Integer Auto-Regressive process of order one, INAR(1). They have shown that the aggregated series follows an Integer Auto-Regressive Moving Average process of order one, INARMA (1,1). Mohammadipour and Boylan [28] have studied theoretically the effects of overlapping temporal aggregation of INARMA processes. They showed that

¹ An important assumption in our analysis is that we start with data that are as disaggregate as our required forecasting output. However, and as one of the referees correctly pointed out the degree of aggregation of the forecasting output does not necessarily need to match with the existing data structure (which may be more aggregate or more disaggregate than the forecasts driving decision making). The degree of aggregation of the forecasting output (i.e. the forecast we use to make decisions) is actually a function of the decision making problem forecasting tries to support. On the contrary inputs to the forecasting process are very often driven by existing data structures. Although the two may indeed match sometimes, this is not always the case.

² <http://www.neural-forecasting-competition.com/NN5/>

the aggregation of an INARMA process over a given horizon results in an INARMA process as well. The conditional mean of the aggregated process was derived as a basis for forecasting. A simulation experiment was conducted to assess the accuracy of the forecasts produced using the conditional mean of the aggregation approach for three INARMA processes: INAR(1), INMA(1) and INARMA(1,1), against that of the non-aggregation approach. The simulation results showed that, in most cases, the aggregation approach provides forecasts with smaller MSEs than non-aggregation ones. The performance of these forecasts was also tested by using two empirical datasets. The first one was from the Royal Air Force (RAF, UK) and consisted of the individual demand histories of 16,000 SKUs over a period of 6 years (monthly observations). The second data set consisted of the demand history of 3,000 SKUs from the automotive industry (over a period of 24 months). The outcome of the empirical investigation confirmed the simulation results.

Willemain et al. [29] have empirically explored the effects of temporal aggregation on forecasting intermittent demand considering the application of Croston's method that has been specifically developed for such demand patterns [30]. The researchers considered 16 empirical data sets of 905 daily observations; the aggregation level was considered to be a week. Results were reported by considering the *MAPE* and the researchers showed a significant reduction in forecasting errors when weekly demand aggregated data were used instead of daily data.

Nikolopoulos et al. [6] have empirically analyzed the effects of non-overlapping temporal aggregation on forecasting intermittent demand requirements. Their proposed approach, called Aggregate-Disaggregate Intermittent Demand Approach (ADIDA), was assessed on 5,000 SKUs containing 7 years history (84 monthly demand observations) from the Royal Air Force (RAF, UK), by means of employing three methods: Naïve, Croston and Syntetos-Boylan Approximation (SBA)[2]. The aggregation level was varied from 2 to 24 months. Comparisons were performed at the original series level (disaggregated demand) and the results showed that the proposed ADIDA methodology may indeed offer considerable improvements in terms of forecast accuracy. The main conclusions of this study were: (1) the ADIDA may be perceived as an important method *self-improving mechanism*; (2) an optimal aggregation level may exist either at the individual series level or across series; (3) setting the aggregation level equal to the lead time length plus one review period $L+1$ (which is the time bucket required for periodic stock control applications) shows very promising results. Spithourakis et al.[31] extended the application of the ADIDA approach to fast-moving demand data. The method's performance was tested on 1,428 monthly time series of the M3-Competition by using the Naïve, SES, Theta, Holt and damped forecasting methods. The empirical results confirmed the previous findings reported by Nikolopoulos et al.[6].

Finally, Babai et al. [9] have also extended the study discussed above [6] by means of considering the inventory implications of the ADIDA framework through a periodic order-up-to-level stock control policy. Three forecasting methods, SES, Croston and SBA were used and the demand was assumed to be negative binomially distributed. Performance was reported through the inventory holding and backlog volumes and costs, for three possible targets Cycle Service Levels (CSL): 90%, 95% and 99%. For high CSLs, the aggregation approach has been shown to be more efficient but for low CSLs it was outperformed by the classical one when Croston's method was used. For SES, the aggregation approach outperforms the classical approach even for low CSLs. The researchers concluded that a simple technique such as temporal aggregation can be as effective as complex mathematical intermittent forecasting approaches.

Before we close this section it is important to note that our working framework resembles to that employed by Widiarta et al. [32, 8, 33] in that the same demand processes and estimator are assumed albeit under a different approach to aggregation. Widiatra et al. examined analytically and by means of simulation the relative performance of some cross-sectional aggregation strategies when demand follows an AR(1) and MA(1) process and when SES is used to extrapolate future demand requirements. Our research focuses on temporal aggregation yet some of the analytical results presented by these researchers have been found to be of relevance and they are utilized in our theoretical developments too. The interface between, and the scope for combining, temporal and cross-sectional aggregation is an issue that has received minimal attention in the academic literature but one that is to be considered in the next steps of our research (please refer also to the last section of the paper).

3. NOTATION AND ASSUMPTIONS

For the remainder of the paper, we denote by:

m : Aggregation level, i.e. number of periods considered to build the block of aggregated demand.

n : total number of periods available in the demand history.

t : Time unit in the original non-aggregated time series. $t=1,2,\dots,n$.

T : Time unit in the aggregated time series. $T=1,2,\dots, \lceil n/m \rceil$.

d_t : Non-aggregated demand in period t

D_T : Aggregated demand in period T

ε_t : Independent random variables for non-aggregated demand in period t , normally distributed with zero mean and variance σ^2

ε'_T : Independent random variables for aggregated demand in period T , normally distributed with zero mean and variance σ'^2

f_t : Forecast of non-aggregated demand in period t , the forecast produced in $t-1$ for the demand in t .

F_T : Forecast of aggregated demand in period T , the forecast produced in $T-1$ for the demand in T .

α : Smoothing constant used in Single Exponential Smoothing method before aggregation, $0 < \alpha \leq 1$

β : Smoothing constant used in Single Exponential Smoothing method after aggregation, $0 < \beta \leq 1$

MSE_{BA} : Theoretical Mean Squared Error (MSE) before aggregation

MSE_{AA} : Theoretical Mean Squared Error (MSE) after aggregation

γ_k : Covariance of lag k of non-aggregated demand, $\gamma_k = Cov(d_t, d_{t-k})$

γ'_k : Covariance of lag k of aggregated demand, $\gamma'_k = Cov(D_T, D_{T-k})$

ϕ : Autoregressive parameter before aggregation, $|\phi| < 1$

ϕ' : Autoregressive parameter after aggregation, $|\phi'| < 1$

θ : Moving average parameter before aggregation, $|\theta| < 1$

θ' : Moving average parameter after aggregation, $|\theta'| < 1$

μ : Expected value of non-aggregated demand in any time period

μ' : Expected value of aggregated demand in any time period

We assume that the non-aggregated demand series d_t follows either a first order moving average, MA(1) or a first order autoregressive data generation process (DGP), AR(1) that can be mathematically written in period t by (1) and (2) respectively.

$$d_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}, \text{ where } |\theta| < 1, \quad (1)$$

$$d_t = \mu(1 - \phi) + \varepsilon_t + \phi d_{t-1}, \text{ where } |\phi| < 1. \quad (2)$$

The m periods non-overlapping aggregated demand D_T can be expressed as a function of the non-aggregated demand series as follows

$$D_{T-k} = \sum_{l=1}^m d_{t-(k-1)m-l} \quad k=1,2,\dots \quad (3)$$

The forecasting method considered in this study is the Single Exponential Smoothing (SES); this method is being applied in very many companies and most of the managers use this method in a

production planning environment due to its simplicity [11]. Using SES, the forecast of demand in period t produced at the end of period $t-1$ is

$$f_t = \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} d_{t-k}, \quad (4)$$

We further assume that the standard deviation of the error term in (1) and (2) above is significantly smaller than the expected value of the demand, so should demand be generated the probability of a negative value is negligible. Constraining ϕ and θ to lie between -1 and 1 in (1) and (2), means that the process is stationary and invertible.

4. THEORETICAL ANALYSIS

In this section we derive the *MSE* of the forecasts generated by considering the non-aggregated and the aggregated demand. Comparisons are to be performed at the original disaggregated level; to that end, the aggregation approach works as follows: firstly buckets of aggregated demand are created based on the aggregation level; then *SES* is applied to these aggregated data and finally the aggregated forecasts are disaggregated by dividing by m , to produce forecasts at the original level. Other disaggregation mechanisms could have been considered [6] but the one employed for the purposes of our research is viewed as realistic from a practitioner perspective and a reasonable approach when dealing with stationary demands. Note that in order to ensure that the forecasting horizon is the same in both the aggregate and the disaggregate cases, the aggregate *SES* forecast is updated in each period when we rebuild the aggregate series.

The comparisons will result in the development of theoretical rules that indicate under which conditions forecasting of aggregated demand is theoretically expected to perform better than forecasting of the non-aggregated demand. These theoretical rules are a function of the aggregation level and the control parameters. The cut-off values to be assigned to the parameters will be the outcome of a numerical analysis to be conducted based on the theoretical results. Having obtained the cut-off values, we can then specify regions of superior performance of the aggregation approach over the non-aggregation one.

In this study the *MSE* is used as a forecast accuracy measure as it is the only theoretically tractable such measure. The *MSE* is similar to the variance of the forecast errors (which consists of the variance of the estimates produced by the forecasting method under concern and the variance of the actual demand) but not quite the same since any potential bias of the estimates may also be taken into account [34]. Since *SES* provides unbiased estimates³ (due to the stationarity of the time series considered in our work) the variance of forecast errors is equal to the *MSE*, i.e. $MSE = Var(\text{Forecast Error})$.

For each process under consideration we calculate the ratio of the *MSE* before aggregation (MSE_{BA}) to the *MSE* after aggregation (MSE_{AA}). A ratio that is lower than one implies that the aggregation approach does not add any value. Conversely, if the ratio is greater than one aggregation performs better than the classical approach.

4.1. MSE Before Aggregation, MSE_{BA}

We begin the analysis by deriving the MSE_{BA} for the MA(1) and AR(1) process. As discussed above the MSE_{BA} is

$$MSE_{BA} = Var(d_t - f_t) = Var(d_t) + Var(f_t) - 2Cov(d_t, f_t), \quad (5)$$

³ Obviously other forecasting methods may also provide unbiased estimates under the stationary demand processes considered in this paper but those are not further considered as their analysis is beyond the scope of this research work.

Subsequently, the three parts of (5) need to be determined: i) variance of demand, ii) variance of the forecast, and iii) the covariance between the demand and the forecast.

4.1.1. MA(1)

When the demand follows an MA(1) process, the autocovariance and autocorrelation functions are [35]:

$$\gamma_k = \begin{cases} (1 + \theta^2)\sigma^2 & k = 0 \\ -\theta\sigma^2 & |k| = 1, \\ 0 & |k| > 1 \end{cases} \quad (6)$$

We begin the evaluation of MSE_{BA} by defining the covariance between the demand and the forecast as follows:

$$\begin{aligned} Cov(d_t, f_t) &= Cov(d_t, \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} d_{t-k}) = \alpha Cov(d_t, \sum_{k=1}^{\infty} (1-\alpha)^{k-1} d_{t-k}) = \\ &= \alpha (Cov(d_t, d_{t-1}) + (1-\alpha)Cov(d_t, d_{t-2}) + (1-\alpha)^2 Cov(d_t, d_{t-2}) + \dots). \end{aligned} \quad (7)$$

Considering that $Cov(d_t, d_{t-k}) = 0$ for all $k > 1$, and substituting (6) into (7) we have:

$$Cov(d_t, f_t) = \alpha\gamma_1. \quad (8)$$

The variance of the forecast is as follows:

$$Var(f_t) = Var(\alpha d_{t-1} + (1-\alpha)f_{t-1}) = \alpha^2 Var(d_{t-1}) + (1-\alpha)^2 Var(f_{t-1}) + 2\alpha(1-\alpha)Cov(d_{t-1}, f_{t-1}). \quad (9)$$

By considering that $Var(f_t) = Var(f_{t-k})$ and $Cov(d_t, f_t) = Cov(d_{t-k}, f_{t-k})$ for all k and substituting (6) and (8) into (9) we get

$$Var(f_t) = \frac{\alpha\gamma_0 + 2\alpha(1-\alpha)\gamma_1}{2-\alpha}, \quad (10)$$

And finally, by using the fact that $\gamma_0 = Var(d_{t-k})$ and substituting (6), (8) and (10) into (5) it is easy to show that:

$$MSE_{BA} = \frac{\gamma_0 - \alpha\gamma_1}{1 - 0.5\alpha}. \quad (11)$$

4.1.2. AR(1)

When demand follows an AR(1) process the following properties hold [35]:

$$\gamma_k = \begin{cases} \sigma^2 & k = 0 \\ \phi^k \gamma_0 & k \geq 1 \end{cases}, \quad (12)$$

To evaluate (5) for AR(1), we begin by deriving the covariance between the forecast and the demand in period t , by substituting (12) into (7), we get:

$$Cov(d_t, f_t) = \frac{\alpha\phi\gamma_0}{(1-\phi+\alpha\phi)}. \quad (13)$$

The variance of forecasts in period t is derived as follows, since $Var(f_t) = Var(f_{t-1})$ and by substituting (12) and (13) into (9), we get:

$$Var(f_t) = \frac{\alpha(1+\phi-\alpha\phi)\gamma_0}{(2-\alpha)(1-\phi+\alpha\phi)} \quad (14)$$

Finally we can calculate the MSE of the forecasts before aggregation. By substituting (12), (13) and (14) into (5), we get

$$MSE_{BA} = \frac{\gamma_0(1-\phi)}{(1-0.5\alpha)(1-\phi+\alpha\phi)} \quad (15)$$

4.2. MSE After Aggregation, MSE_{AA}

In this section we proceed with the derivation of the MSE of the forecasts for the aggregation approach. Demand is first aggregated to produce high frequency demand forecasts based on SES followed by the disaggregation of such forecasts to produce one-step-ahead estimates at the original level by dividing the aggregated forecast by the aggregation level m . The MSE_{AA} is defined as

$$MSE_{AA} = Var\left(d_t - \frac{F_T}{m}\right) = Var(d_t) + \frac{1}{m^2} Var(F_T) - \frac{2}{m} Cov(d_t, F_T) = \gamma_0 + \frac{1}{m^2} Var(F_T) - \frac{2}{m} Cov(d_t, F_T), \quad (16)$$

By applying SES, the aggregated forecast for period T is defined as

$$F_T = \sum_{k=1}^{\infty} \beta(1-\beta)^{k-1} D_{T-k}. \quad (17)$$

4.2.1. MA(1)

When the non-aggregated series follows an MA(1) process, the aggregated series also follows an MA(1) process but with a different parameter value [15, 36].

Based on Wei [35] we can show that the relationship between the autocovariance function of the non-aggregated and the aggregated demand is

$$\begin{cases} \gamma'_0 = m\gamma_0 + 2(m-1)\gamma_1 \\ \gamma'_1 = \gamma_1 \end{cases}. \quad (18)$$

From (A-2) and (B-5) in Appendix A and B respectively, we have:

$$Cov(d_t, F_T) = \beta\gamma_1, \quad (19)$$

$$\text{Var}(F_T) = \frac{\beta\gamma'_0 + 2\beta(1-\beta)\gamma'_1}{2-\beta}, \quad (20)$$

So, by substituting (19) and (20) into (16) and then using (18) into that, the *MSE* of the forecast after aggregation is

$$\text{MSE}_{AA} = \left(\gamma_0 + \frac{1}{m^2} \left(\frac{\beta m \gamma_0 + 2m\beta\gamma_1 - 2\beta^2\gamma_1}{2-\beta} \right) - \frac{2\beta\gamma_1}{m} \right). \quad (21)$$

4.2.2. AR(1)

If the non-aggregated series follows an AR(1) process then the aggregated series follows an ARMA(1,1) process [14, 36]. It can be shown that the following properties hold when the process is ARMA(1,1):

$$D_T = \mu'(1-\phi') + \varepsilon'_T + \phi' D'_{T-1} - \theta' \varepsilon'_{T-1}, \text{ where } |\theta'| < 1, |\phi'| < 1, \quad (22)$$

$$\gamma'_k = \begin{cases} \frac{1-2\phi'\theta' + \theta'^2}{1-\phi'^2} \sigma'^2 & k=0 \\ \frac{(\phi' - \theta')(1-\phi'\theta')}{1-\phi'^2} & |k|=1 \\ \phi' \gamma'_{k-1} = \phi'^{k-1} \gamma'_1 & |k| > 1 \end{cases} \quad (23)$$

From Wei [35] we can show the relation of autocovariance function of non-aggregated and aggregate demand series when the non-aggregated series follows an AR(1) process:

$$\gamma'_0 = \gamma_0 \left(m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right), \quad (24)$$

$$\gamma'_1 = \gamma_0 \left(\sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right), \quad (25)$$

$$\phi' = \phi^m, \quad (26)$$

and for all $k > 1$, we have:

$$\gamma'_k = \gamma_0 \left(\phi^{(k-1)m+1} + 2\phi^{(k-1)m+2} + \dots + m\phi^{km} + (m-1)\phi^{km+1} + \dots + \phi^{(k+1)m-1} \right). \quad (27)$$

In Appendices A and B (equations (A-3) and (B-6)) we show that:

$$\text{Cov}(d_t, F_T) = \frac{\beta\phi\gamma_0}{1-\phi^m + \beta\phi^m} \times \frac{1-\phi^m}{1-\phi}, \quad (28)$$

$$\text{Var}(F_T) = \frac{\beta\gamma'_0}{2-\beta} + \frac{2\beta(1-\beta)\gamma'_1}{(2-\beta)(1-\phi' + \beta\phi')}, \quad (29)$$

Considering equations (24), (25), (26), (28) and (29), the MSE of the forecasts at the disaggregated level is given by:

$$MSE_{AA} = \gamma_0 \left(1 + \frac{1}{m^2} \left(\frac{\beta \left(m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right)}{2-\beta} + \frac{2\beta(1-\beta) \left(\sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right)}{(2-\beta)(1-\phi^m + \beta\phi^m)} \right) - \frac{2}{m} \left(\frac{\beta\phi}{1-\phi^m + \beta\phi^m} \times \frac{1-\phi^m}{1-\phi} \right) \right). \quad (30)$$

5. MSE COMPARISON RESULTS

The effectiveness of temporal aggregation as compared to non-aggregation may be assessed by analyzing the ratio of their MSE s. Recall from Section 4, that a value of MSE_{BA}/MSE_{AA} greater than 1 implies that the aggregation approach is superior to the non-aggregation one, whereas a value that is lower than 1 implies the opposite. A ratio value equals to 1 means that performance is the same.

In subsection 5.1, we investigate the impact of the aggregation level, m , the smoothing constants α and β and the moving average θ or the autoregressive parameter ϕ on the ratio of MSE_{BA}/MSE_{AA} by varying their values. In subsection 5.2, we determine analytically the conditions under which one approach outperforms the other. Finally in subsection 5.3 we are concerned with the determination of the optimum aggregation level.

5.1. Impact of the Parameters – Sensitivity Analysis

In this section the effect of the parameters m, α, β, θ and ϕ on the ratio MSE_{BA}/MSE_{AA} is analyzed. Note that m, α, β , are control parameters often set by the forecaster, whereas θ and ϕ are parameters associated with the underlying demand generation process (process parameters). Therefore, we are interested to know which values of the control parameters lead to a ratio higher than 1, for any given values of the process parameters.

In real world settings data could typically be aggregated as weekly ($m=7$) from daily data, yearly ($m=4$) from quarterly, monthly ($m=4$) from weekly, quarterly ($m=3$) from monthly, semi-annually ($m=6$) from monthly and annually ($m=12$) from monthly data or it may also be aggregated at some other level to reflect relevant business concerns (e.g. equal to the lead time length).

Given the considerable number of control parameter combinations it is natural that only some results may be presented here. The simulation output was judged to be represented sufficiently through the consideration of $\alpha = 0.1, \alpha = 0.5, m=2, m=12$ and the whole range of β, θ and ϕ .

5.1.1. MA(1)

Figure 1 presents the impact of the control parameter β on the ratio of MSE_{BA}/MSE_{AA} for $m = 2, 12$ and $\alpha = 0.1, 0.5$, when the non-aggregated demand series follows an MA(1) process. Shaded areas represent a behavior in favor of the non-aggregation approach. The results show that for a fixed value of α , by increasing the aggregation level, the aggregation approach provides more accurate forecasts than the non-aggregation one. On the other hand, when considering a fixed value of the aggregation level, increasing β results in a deterioration of the aggregation approach. If the selected smoothing constant value after aggregation, β , is considerably higher than the smoothing constant used at the original data, α , then the aggregation approach is not preferable. Alternatively, the aggregation approach may produce more accurate forecast unless θ takes highly negative values.

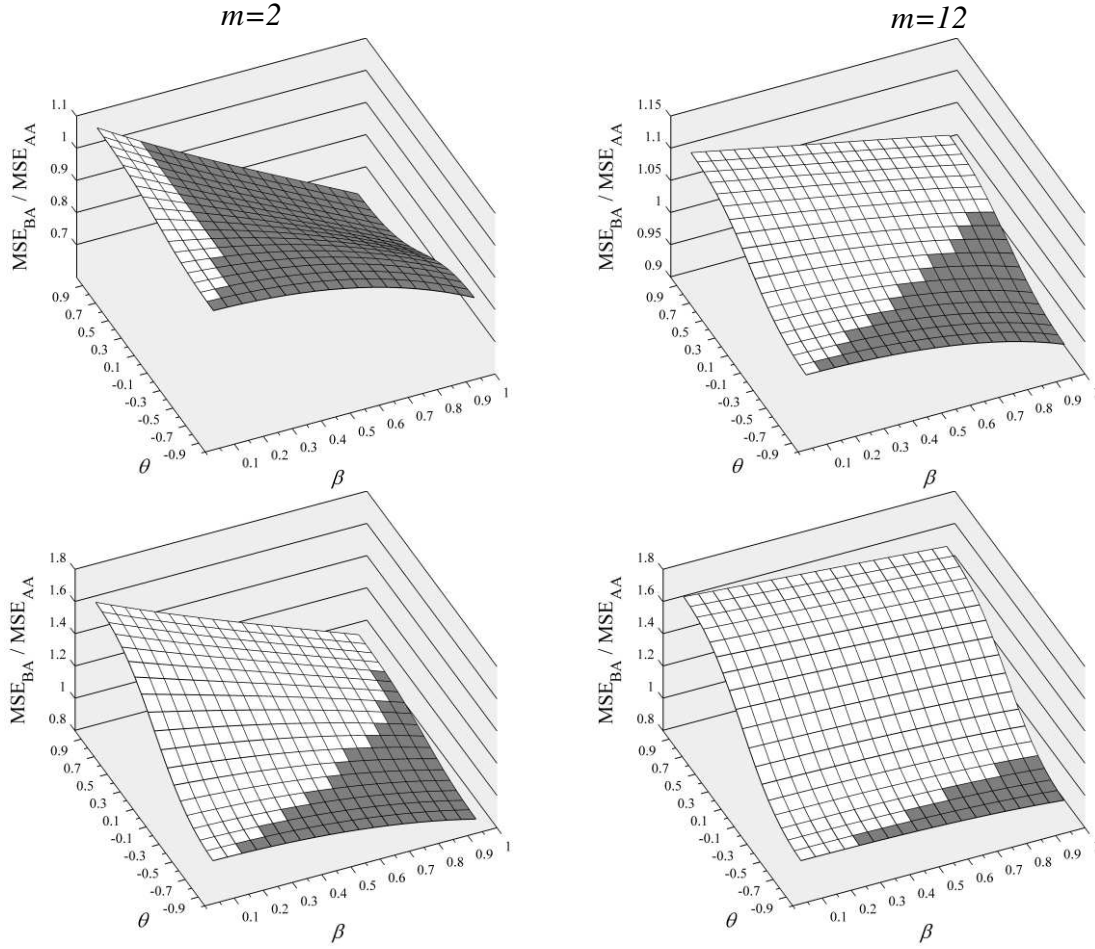


Figure 1. Impact of m , θ , α and β on the MSE ratio for $\alpha = 0.1$ (top) and $\alpha = 0.5$ (bottom)

In the particular case where the smoothing constant parameters before and after aggregation are identical ($\alpha = \beta$), the aggregation approach outperforms the non-aggregation one in all cases, except those associated with high negative values of θ (high positive autocorrelation). Moreover, even in those cases, when increasing the aggregation level the performance of the aggregation approach is improved.

The impact of the smoothing parameter β and the aggregation level m is quite intuitive. In fact, it is obvious that the coefficient of variation (CV) of the non-overlapping temporally aggregated demand is smaller than the CV of the original (non-aggregated demand) and we can easily show that by increasing the aggregation level the coefficient of variation of demand is further reduced. This means that high aggregated order series are associated with less dispersion than low aggregated order series. In addition, by considering the autocovariance function before and after aggregation, we may show that the application of non-overlapping temporal aggregation decreases the value of the autocorrelation function and increasing the aggregation level leads to a higher reduction in the autocorrelation which eventually becomes zero. That is, the aggregated series tend towards a white noise process in which case small values of the smoothing constant lead to smaller $MSEs$. Therefore, setting β to be small (β should be smaller than α) in conjunction with high aggregation levels provides an advantage to the aggregation approach; this is confirmed by the results presented in Figure 1.

We note that even if the selected β is smaller than α , there are cases in which the aggregation approach is not preferable. This can be attributed to the potential high positive autocorrelation between demand periods. For negative values of θ , the autocorrelation is positive; for positive values of θ the autocorrelation is negative and for white noise process, the autocorrelation is zero. Aggregation of highly positively correlated series doesn't add as much value as aggregating series with less positive autocorrelation.

These examples show that the performance superiority of each approach is a function of all the control and the process parameters and the selection of the control parameters α , β and m , influence the effectiveness of the aggregation approach in conjunction with the consideration of the process parameters. In subsection 5.2.1 we attempt to identify the conditions under which each approach produces more accurate forecasts for a fixed value of α .

5.1.2. AR(1)

Figure 2 presents the impact of control parameters m , α , β on the ratio of MSE_{BA}/MSE_{AA} for $m = 2, 12$ and $\alpha = 0.1, 0.5$, when the non-aggregated demand series follows an AR(1) process. As in the case of the MA(1) process it is easy to see that the superiority of each approach is a function of all the control parameters. The results show that for a fixed value of α , increasing the aggregation level results in accuracy improvements of the aggregation approach. Conversely, for a fixed aggregation level, increasing β results in a deterioration of the performance. In addition, β should be generally smaller than α in order for the aggregation approach to produce more accurate forecasts.

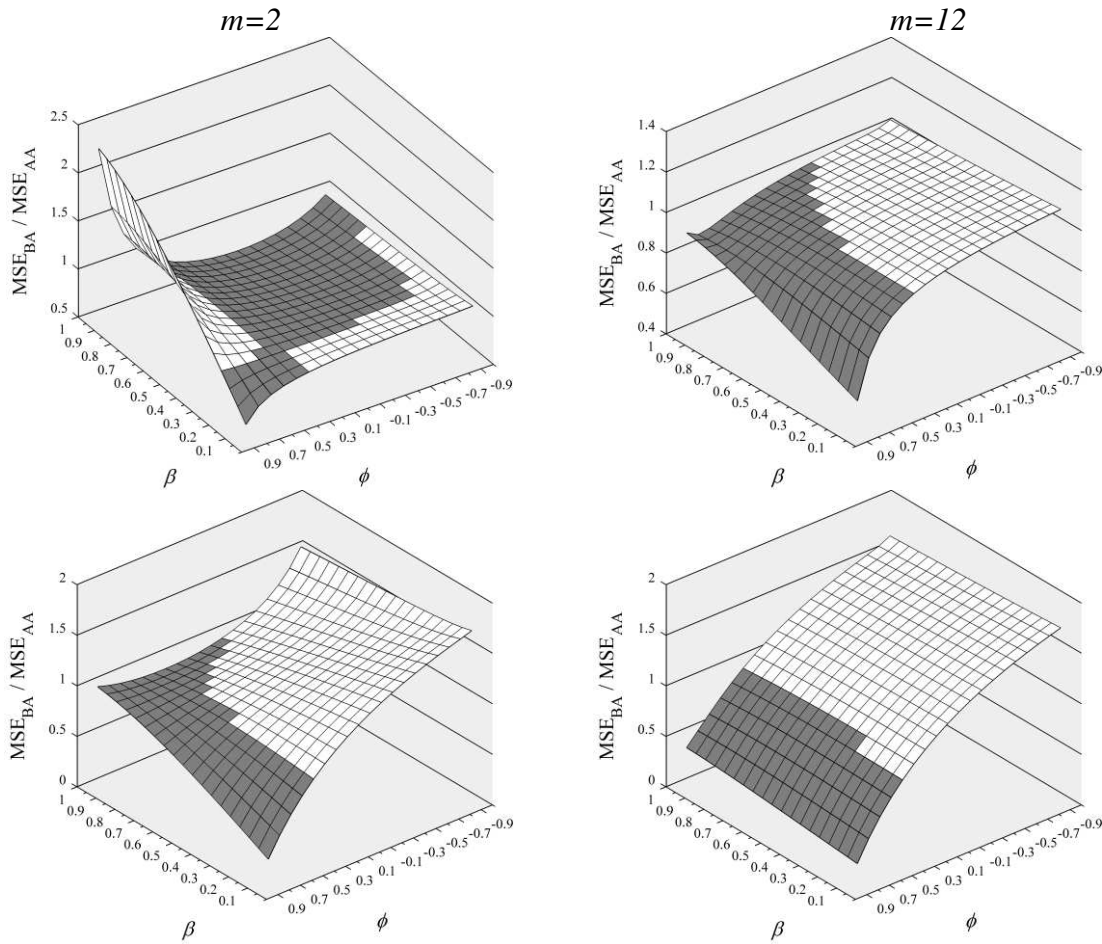


Figure 2. Impact of m , ϕ , α and β on the MSE ratio for $\alpha = 0.1$ (top) and $\alpha = 0.5$ (bottom)

As in the case of the MA(1) process, when the smoothing constant parameters before and after aggregation are identical (i.e. $\alpha = \beta$), the aggregation approach outperforms the non-aggregation one in all cases, except those that are associated with highly positive autocorrelation (highly positive values of ϕ). In those exceptional cases the comparative performance of the two approaches is insensitive to the increase of the aggregation level. The impact of the smoothing parameter β and the aggregation level

m on the ratio is similar to that reported for the MA(1) process. Figure 2 shows that for highly positive values of the autoregressive parameter ϕ the aggregation approach does not work well and the non-aggregation approach provides more accurate results. This is generally true regardless of the values employed by the other control parameters. Therefore, the aggregation approach is not recommended in such cases. When ϕ is positive, the series is 'slowly changing' or can be considered as a positively autocorrelated process. In addition when the non-aggregated demand follows an AR(1) process, the autocorrelation spans all time lags (not only lag 1). As such, for highly positive ϕ values no level of aggregation may improve the accuracy of forecasts.

Hence, when the non-aggregated demand follows an AR(1) process, the aggregation approach may lead to accuracy improvements when the aggregation level, m , is high and the smoothing constant after aggregation β is small. However, for highly positive values of the autoregressive parameter ϕ , the aggregation approach is not recommended (especially when β is bigger than α).

In summary, what may be concluded at the end of this subsection is that for both the MA(1) and AR(1) processes, if the demand data is positively autocorrelated then the non-aggregation approach works better than the aggregation one; in those cases the non-aggregation approach exploits better the very important recent information (i.e. d_t) (though it is more prone to noise). On the contrary, when the recent demand information is not that crucial then a longer term view of the demand is preferable (if one properly selects how to use long term demand information through m and β). Moreover, the aggregation performance under the MA(1) and AR(1) process is slightly different due to the nature of these processes. Positive autocorrelation under an AR(1) process, with a maximum value equal to +1, is potentially higher than that associated with an MA(1) process (with a maximum value equal to 0.5) which renders the range of outperformance of the non-aggregation approach larger under the AR(1) process. In subsection 5.2.2 we theoretically determine the conditions under which each approach outperforms the other one under the AR(1) process.

5.2. Comparative Performance

Having conducted a sensitivity analysis in subsection 5.1, we now identify analytically the conditions under which each approach outperforms the other one.

5.2.1. MA(1)

The ratio of the MSE_{BA} to MSE_{AA} when the non-aggregated demand follows an MA(1) process is a function of the moving average parameter, the smoothing constant before and after aggregation (α and β) and the aggregation level. Considering that the aggregation level may only get integer values greater than or equal to two, we wish to determine the value β that enables the aggregation approach to perform better. Here we consider the entire range of possible values for β but the smoothing constant is a parameter that is set to its optimal value by practitioners, normally by minimizing the MSE . From (11) it is obvious that MSE_{BA} is monotonically increasing in α , as the derivative of MSE_{BA} is positive for all values of θ in $[-1,1]$. Hence, MSE_{BA} can be minimized by having the smallest possible value of α , which makes sense for a stationary process. However, we do wish to note that in our theoretical analysis we disregard the issue of initialization of the forecasting process. This is an important issue to be mentioned (since with very low α values a bad initialization implies inaccurate estimates of the future demand as the forecast will basically be kept constant) but one that is not considered as part of our research.

To show the conditions under which the aggregation approach outperforms the non-aggregation approach, we need to set $MSE_{BA}/MSE_{AA} > 1$. From this inequality we can obtain the following result:

THEOREM 1: If the time series of the non-aggregated demand follows an MA(1) process, then:

- If $\beta < \beta_1$, the aggregation approach provides more accurate forecasts.
- If $\beta = \beta_1$, both strategies perform equally.
- Otherwise, the non-aggregation approach works better.

where

$$\beta_1 = \frac{\left(-(m^2\eta + m(1 + \theta^2) + 2m\theta) + \sqrt{\left((m^2\eta + m(1 + \theta^2) + 2m\theta) \right)^2 + 8(2\theta - 2m\theta)m^2\eta} \right)}{2(2\theta - 2m\theta)}, \quad (31)$$

and

$$\eta = \frac{\alpha(1 + \theta)^2}{2 - \alpha}. \quad (32)$$

PROOF: the proof of Theorem 1 is given in Appendix C.

The results demonstrate that, for given α and m , there always exists a value of β such that aggregation approach outperforms non-aggregation one.

Note that β_1 is always positive, consequently choosing $\beta < \beta_1$ guarantees that the aggregation approach always outperforms the disaggregation one in this region. Hence, the value of β_1 reflects a cut-off point that may be used in practice for the selection of the smoothing constant value to be used for the aggregated series. The cut-off point reflects all the qualitative discussion provided in the previous subsection as to when aggregation outperforms the non-aggregation approach.

5.2.2. AR(1)

We work in a similar way to the MA(1) process by setting the ratio MSE_{BA} to MSE_{AA} greater than 1 in order to identify the conditions under which the aggregation approach performs better. These conditions are summarized by the selection procedure presented in Appendix D.

As we mentioned earlier the smoothing constant is often set by practitioners to its optimal value, so it is more interesting to discuss in our work the cases where such a value is considered. To do so, we first find the value that minimizes the MSE_{BA} and then we find the value of the smoothing constant after aggregation that leads to more accurate forecasts. The optimal value that minimizes (15), denoted by α^* , obtained by solving the first derivative of (15), is given by (see [32]):

$$\alpha^* = \begin{cases} (3\phi - 1)/2\phi & 1/3 < \phi < 1 \\ \omega & -1 < \phi \leq 1/3 \end{cases}. \quad (33)$$

where $\omega > 0$ is a very small positive value.

By considering the optimal value of the smoothing constant before aggregation, we have two different cases. From $MSE_{BA}/MSE_{AA} > 1$ and (33) we can get the following results.

Case 1. $1/3 < \phi < 1$. In this case, $\alpha^* = (3\phi - 1)/2\phi$

THEOREM 2: If the time series of the non-aggregated demand follows an AR(1) process, where $1/3 < \phi < 1$ and the optimal smoothing constant, $\alpha^* = (3\phi - 1)/2\phi$, is used to determine the non-aggregated demand forecast, then the non-aggregation approach always provides more accurate forecast

than the aggregation one, regardless of the smoothing constant parameter after aggregation, β , and the aggregation level, m .

PROOF: the proof of Theorem 2 is given in Appendix E.

Case 2. $-1 < \phi \leq 1/3$. In this case α^* is a very small positive number.

THEOREM 3. If the time series of the non-aggregated demand follows an AR(1) process, where $-1 < \phi \leq 1/3$ and the optimal smoothing constant used to determine the non-aggregated demand forecast, $\alpha^* < 0.05$, then:

- If $\beta < \beta_1$ the aggregation approach provides more accurate forecast.
- If $\beta = \beta_1$ both strategies perform equally.
- Otherwise, the non-aggregation approach works better.

where

$$\beta_1 = \frac{-\left((1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\right) + \sqrt{\Delta}}{2(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)}. \quad (34)$$

(ξ_1, ξ_2, η and Δ are given in Appendix D)

PROOF: The proof of Theorem 3 is given in Appendix E.

As for the MA(1) process, the above results provide a cut-off point that may be used in practice for the selection of the smoothing constant in order to obtain an outperformance of the aggregation approach when AR(1) processes are considered. Obviously, as the cut-off point increases for high aggregation levels, it is clear that this implies a considerable range of the smoothing constant of the aggregated series where there is a benefit of using the aggregation approach. Hence, these results provide a comprehensive way of managing the process of forecasting of AR(1) processes when the autoregressive parameter is known and when the intention is to optimize the smoothing constant for the non-aggregated series.

5.3. Optimal Aggregation Level

The objective of this section is to identify the optimal aggregation levels that maximize the ratio or equivalently minimize the MSE_{AA} for each demand process under consideration. To do so, we evaluate the ratio of MSE_{BA} to MSE_{AA} for the whole range of the control parameters.

5.3.1. MA(1)

In order to obtain the optimal aggregation level when the non-aggregated demand series follows an MA(1), we consider the following theorem.

THEOREM 4: If the non-aggregated demand series follows an MA(1) process, then the optimal aggregation level is the highest level in any considered range.

Suppose, aggregation is to be tested in a range $[u_1, u_2]$, where u_1 and u_2 are the lower and upper bound respectively and they are positive integer numbers. The optimal aggregation level will always be u_2 .

PROOF: A calculation of the first derivative of MSE_{AA} with respect to m and a numerical analysis for $m \geq 2$ shows that MSE_{AA} is a decreasing function of m , which means that the ratio MSE_{BA}/MSE_{AA} is an increasing function of m . Therefore, a higher value of the aggregation level results in a higher value of the ratio MSE_{BA}/MSE_{AA} .

5.3.2. AR(1)

In order to obtain the optimal aggregation level when the non-aggregated process follows an AR(1) process, we consider a numerical investigation since from (30) it is obvious that the calculation of the first derivative is infeasible. Two examples have been considered: i) the whole range of ϕ where $\alpha = 0.15$ and $\beta = 0.1$; ii) the case 2 discussed in 5.2.2 with an optimal value of α .

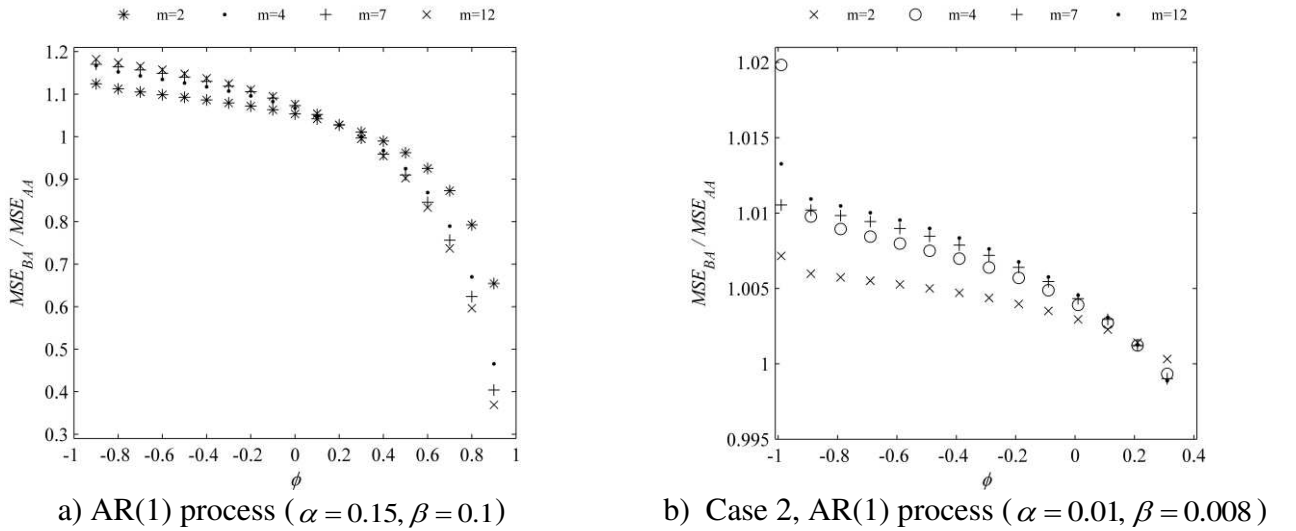


Figure 3. MSE ratio for different values of m

Figure 3a shows that the value of the aggregation level that maximizes the MSE ratio changes when varying the control parameter values. For negative and lower positive values of ϕ , i.e. $-1 < \phi \leq 1/3$, the forecast accuracy of the aggregation approach increases with the aggregation level while for higher positive values of ϕ , i.e. $1/3 < \phi < 1$, this is not true. Let us analyze the two different cases in which we use the optimal smoothing constant value for MSE_{BA} .

Case 1. $1/3 < \phi < 1$. In this case the optimal smoothing constant parameter $\alpha^* = (3\phi - 1)/2\phi$ is used and we show in subsection 5.2.2 that the MSE ratio is always lower than 1.

Case 2. $-1 < \phi \leq 1/3$. In this case a very small smoothing constant value, $\alpha^* \leq 0.05$, is used. The MSE ratio for different aggregation levels is shown in Figure 3b for $-1 < \phi \leq 1/3$ and a numerical example of α and β values where $\beta < \alpha$. This figure shows that the aggregation approach is associated with more accurate results for higher aggregation levels.

6. EMPIRICAL ANALYSIS

In this section we assess the empirical validity of the main theoretical findings of this research. In the following subsection we provide details of the empirical data available for the purposes of our investigation along with the experimental structure employed in our work. In subsection 6.2 the actual empirical results are presented.

6.1. Empirical Dataset and Experiment Details

The demand dataset available for the purposes of our research consists of weekly sales data over a period of two years for 1,798 SKUs from a European grocery store. The Time Series Expert Modeling function of SPSS (version 19) has been used to identify the underlying ARIMA demand process for each series and estimate the relevant parameters. It was found that more than 43% of the series may be represented by the processes considered in our research. In particular, 30.3% of the series (544 series) were found to be ARIMA(1,0,0) and 13% (233 series) to be ARIMA(0,0,1). Other popular processes identified were: ARIMA(1,0,1) (8.6%), ARIMA(0,0,0) (16.3%) and ARIMA(0,1,1) (23.7%). This analysis provides some empirical justification on the frequency of stationary, and in particular MA(1) and AR(1), processes in real world practices.

In Table 1 we summarize the characteristics of the SKUs relevant to our study by indicating the estimated parameters for the MA(1) and AR(1) processes. To facilitate a clear presentation, the estimated parameters are grouped in intervals and the corresponding number of SKUs is given for each such interval. The average θ and ϕ value per interval is also presented for the MA(1) and AR(1) process respectively. This categorization allows us to compare the empirical results with the theoretical findings. We must remark that the θ parameter values are all but one negative and the ϕ parameter values are all but one positive. As such, the data do not cover the entire theoretically feasible range of the parameters. Some studies [37, 38, 39] that have considered empirical AR(1) processes, have reported that is common to have positive correlation/high value of autoregressive parameters in the consumer product industries which is also the case in the dataset used in our research. Replication of our findings in bigger datasets is certainly an avenue for further research in this area and this issue is discussed in more detail in the next section of the paper.

Table 1. Processes Present in the Empirical Data Set

a) MA(1)			b) AR(1)		
θ intervals	Average of θ	No. of SKUs	ϕ intervals	Average of ϕ	No. of SKUs
$[-.8, -.7[$	-0.7252	1	$[-.1, 0[$	-0.2240	1
$[-.7, -.6[$	-0.6329	9	$[.1, .2[$	0.1981	2
$[-.6, -.5[$	-0.5393	39	$[.2, .3[$	0.2534	84
$[-.5, -.4[$	-0.4471	72	$[.3, .4[$	0.3549	125
$[-.4, -.3[$	-0.3509	57	$[.4, .5[$	0.4479	127
$[-.3, -.2[$	-0.2520	48	$[.5, .6[$	0.5512	121
$[-.2, -.1[$	-0.1989	6	$[.6, .7[$	0.6433	63
$[0, 1[$	0.2831	1	$[.7, .8[$	0.7352	18
			$[.8, .9[$	0.8256	3
Total number of SKUs:		233	Total number of SKUs:		544

The data series have been divided into two parts. The first part (within sample) consists of 62 time periods and is used in order to initialize the SES estimates. The second part consists of the remaining 41 time periods and is used for the evaluation of the performance (out-of-sample).

The values of the smoothing constants were varied from 0.05 to 0.95 with a step increase of 0.05. In the classical (non-aggregate) approach, we first calculate 41 one-step ahead forecasts for each series and then we calculate the variance of the forecast error.

In order to obtain the forecasts via the aggregation approach we start by creating non-overlapping buckets of aggregated data based on a specified aggregation level and then we apply SES to these aggregated data.

Aggregation level = 2: Starting from the 62nd weekly observation in the initial (within sample) part, we sum observations backwards in buckets of two (2), resulting in a bi-weekly series consisting of 31 aggregated observations. The average of aggregated series is obtained and is used as the SES forecast for the first bucketed period 1. SES is then applied all the way up to producing a forecast for bucket 32 which is then divided by 2 (the aggregation level, $m=2$) and it gives a forecast for periods 63 and 64.

We drop the forecast for period 64 and record the one for 63 (they are equal anyway). Then we start creating buckets of 2 periods from period 63 backwards. So we create another 31 buckets and the very first observation (period 1 in the original data) is not used anymore. We average these buckets (they are different from those created before), we use that average as the SES forecast for the first bucket, we continue using SES until the point that we produce a forecast for bucket 32 (periods 64 and 65). We keep the forecast for period 64 and so on. In the next period we bucket backwards from period 64 ending up with 32 buckets and continue like this until obtain the forecasts for 41 periods ahead.

Aggregation level = 3 . . . 24: Similarly, we continue with time buckets of up to 24 periods. At this point there are 2 aggregated biweekly observations ($2 \times 24 = 48$), so 14 weekly observations at the start of the original series remain unused.

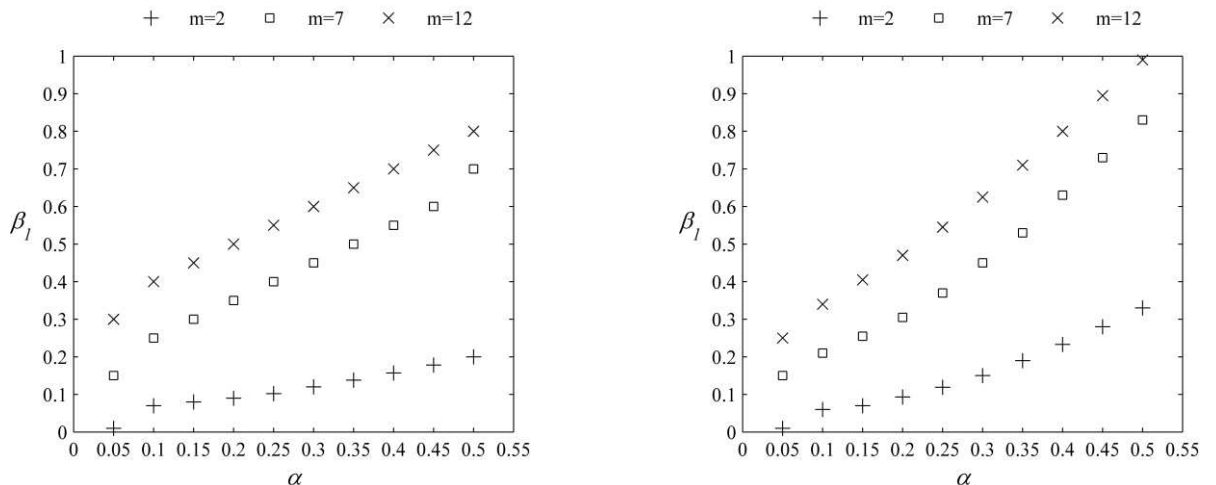
Finally, the value of the variance of the forecast error before aggregation is divided by the variance of the forecast error after aggregation, to obtain the ratio of MSE_{BA} to MSE_{AA} .

6.2. Empirical Results

In Section 5 we examined analytically the conditions under which the aggregated forecasts may perform better than the non-aggregated ones using the ratio of MSE_{BA} to MSE_{AA} . The cut-off points of the smoothing constant of the aggregate series β that should be used (i.e. any value of β that is lower than the cut-off point β_1 implies an outperformance of the aggregation approach) have also been determined for both the MA(1) and AR(1) process. In the following figures we present the results of the empirical analysis and we investigate the degree to which they validate our theoretical findings.

In Figure 4a we show the cut-off point β_1 for fixed values of α and m when the non-aggregated demand of the SKUs follows an MA(1) process. Please recall that the cut-off point β_1 is the value below which any β value implies that the aggregation approach outperforms the non-aggregation one. Note that we only show the results for $\alpha \leq 0.5$ since this range is viewed as realistic for the stationary processes considered in this work.

The empirical results show that for a low aggregation level $m=2$, the cut-off point is relatively low since $\beta_1=0.2$ for a relatively high α value equal to 0.5. In that case, the MSE reduction when $\beta=0.05$ is equal to 8.89% and the MSE ratio decreases for higher values of β . Obviously, the cut-off value considerably increases when the aggregation level increases. For example, when we consider the aggregation level $m=12$, the cut-off point may go up to $\beta_1=0.8$ for α value equal to 0.5. In that case the MSE reduction when $\beta=0.05$ is equal to 12.13%. This shows the considerable region where the aggregation approach outperforms the non-aggregation one for high aggregation levels. Hence, increasing the aggregation level improves the performance of the aggregation approach and the best results can be achieved for small values of β and high aggregation levels m . These empirical results generally confirm the theoretical findings.



a) MA(1) process

b) AR(1) process with $-1 < \phi \leq 0.33$

Figure 4. Cut-off points of β implying an outperformance of the aggregation approach for different values of α and m .

Figure 4b shows the cut-off point β_1 for fixed values of α and m when the SKUs have a non-aggregated demand that follows an AR(1) process with $-1 < \phi \leq 0.33$. The empirical results show that for a low aggregation level $m=2$, low β values should be selected in order to have an outperformance of the aggregation approach. For example when we use an aggregation level $m=2$, the cut-off point $\beta_1=0.33$ for an α value equal to 0.5 and the MSE reduction when $\beta=0.05$ is equal to 12.45%. The cut-off points considerably increase when the aggregation level increases. Figure 4b shows also that for an α value equal to 0.5 and when the aggregation level $m=12$, the cut-off point β_1 is almost equal to 1, which means that the aggregation approach always outperforms the non-aggregation one in that case. That results also in a MSE reduction equal to 15.11% that decreases for higher values of β . However, it should be noted that for the SKUs where $0.33 < \phi < 1$, the empirical results show that when the optimal value of α is used for all values of β and m , the non-aggregation approach outperforms the aggregation one.

The empirical analysis confirms overall the results of the theoretical evaluation both for the MA(1) and AR(1) processes. What can be concluded here is that there is a considerable range of the values of the smoothing constant of the aggregated series that implies a benefit of using the aggregation approach. This benefit can also be substantial for high aggregation levels and low smoothing constants. Note that such analysis can be utilized as an indicator on when the aggregation approach should be used and which parameters lead to the outperformance of this approach.

7. IMPLICATIONS, CONCLUSIONS AND FURTHER RESEARCH

Aggregation is an appealing approach to reduce demand uncertainty for both fast and slow moving items. Moreover, most inventory forecasting software packages support aggregation of data. Although this would typically cover cross-sectional aggregation (i.e. aggregation across series), the consideration of temporal aggregation has been neglected by software manufacturers although it constitutes an opportunity for adding more value to their customers. In this paper we have analytically evaluated the effectiveness of the non-overlapping temporal aggregation approach on forecasting performance when non-aggregate series follow a first order moving average [MA(1)] or a first-order univariate autoregressive [AR(1)] processes. Forecasting was assumed to be relying upon a Single Exponential Smoothing (SES) procedure and the analytical results were complemented by a simulation experiment on theoretically generated data as well as experimentation with an empirical dataset of a European superstore.

Although it is true that the fast changing market environment results in many demand processes being non-stationary in nature, stationary demands may still constitute a realistic assumption. The empirical data available for the purposes of our research confirm such a statement and provide support for the frequency with which both MA(1) and AR(1) processes are encountered in real world applications. We have analytically considered the entire range of possible moving average and autoregressive parameter values. Negative parameter values result in positive (negative) autocorrelation and positive values in negative (positive) autocorrelation under the assumption of a MA(1) (AR(1)) process. Although positive autocorrelation is intuitively more common in real world data (this was also evidenced in our dataset) due to the repeat purchasing behavior of most shoppers, negative autocorrelation can be attributed to a ‘variety seeking’ behavior and can be present in many product categories too. Finally, SES is a most commonly employed forecasting procedure in industry and although its application implies a non-stationary behavior (SES is optimal for an ARIMA(0,1,1) process) the method is unbiased and most often used for stationary demands as well. In summary, we feel that the problem setting we consider is a very realistic one and our work sets the basis for more studies in this area.

Analytical developments were based on the consideration of the Mean Squared Error (MSE) before and after aggregation (MSE_{BA} / MSE_{AA}) and comparisons were undertaken at the original (non-aggregated) demand level. The conditions under which one approach outperforms the other were identified and the main findings can be summarized as follows.

- First of all, the performance of aggregation was generally found to improve as the aggregation level increases. The rate of improvement though is lower for the AR(1) process compared to MA(1). MSE percentage reductions may be as high as 15% with significant potential implications for inventory control since the MSE translates directly to safety stocks. Practitioners should always opt for the highest possible aggregation level. However, it is important to note that consideration of high aggregation levels is subject to data availability. Although this progressively becomes less of an issue in modern business settings, clearly aggregation may not constitute a viable option when short demand histories are available. Tremendous recent developments in terms of computing storage capacity facilitate the accumulation of very lengthy series although we have come across situations/companies where only a few years' data is stored. In such cases aggregation may not be further considered. Long historical data series do not only allow for the more accurate estimation of the series' components but also permit the application of temporal aggregation approaches.
- Second, the performance of aggregation improves as the smoothing constant value employed at the aggregated series reduces. Our analytical results show that as the level of aggregation increases the auto-correlation of the series reduces necessitating the employment of low smoothing constant values. This is an important finding from a practitioner perspective since managers may set such values conveniently low to maximize the benefits derived from the aggregation approach. The smoothing constant value after aggregation should be generally set to be smaller than the smoothing constant before aggregation and specific rules and cut-off points have been offered for making such decisions.
- Third, and following from the above, our analysis suggests that there are shades of aggregation (at one extreme no data aggregation) and shades of responsiveness of the forecast parameters (α , β). Our findings suggest that the dominant solutions are either pure white (disaggregate data and responsive parameters) or pure black (aggregate data and stable forecasting algorithms with low β). This is, up to a certain extent, an expected outcome given the hypothesized stationarity but: i) it is not obvious and to the best of our knowledge has never been shown before; ii) it sheds light to the general trade-off between stable forecast parameters (low smoothing constant values) that filter noise rather effectively but fail to react to changes in demand quickly and responsive forecast parameters (relatively higher smoothing constant values) that however are noise sensitive.
- Fourth, for high levels of positive autocorrelation in the original series the aggregation approach may be outperformed by the non-aggregation one. This is an intuitive finding since at any time the most recent demand information is so precious in that case that the disaggregate process works better as it fully exploits such recent information. However, on the contrary, for low positive autocorrelation when the recent demand information is not that crucial then a more long term view on demand is preferable, which can be obtained as discussed above by selecting high aggregation levels and low smoothing constants. This is also an important empirical insight since managers may know what to expect (in terms of any potential gains) based on the autocorrelation levels present in their series.

Our discussions with practitioners have revealed a misconception that aggregation reduces variability, something that is clearly not the case. Aggregation does though reduce the coefficient of variation leading to lower uncertainty. Practitioners have also expressed concerns with regards to the intuitively appealing loss of information associated with temporal aggregation. However, this concern is conditioned to short demand histories. Should long demand series be available the loss of information resulting from aggregation is outweighed by the benefits of uncertainty reduction.

Given the current under-consideration of temporal aggregation in inventory forecasting software solutions and given its value as a promising uncertainty reduction time series transformation approach that this study has revealed, research into any of the following areas would appear to be merited.

- There is a lack of empirical evidence in the area of temporal aggregation and a great need to expand the current knowledge base. Research on more extensive datasets (as well as analysis of empirical forecasting performance on measures other than the MSE) should allow a better understanding of the difficulties and benefits associated with aggregation.
- Expansion of the analytical work discussed in this paper on higher order stationary processes and more importantly on non-stationary processes is a very important issue both from an academic and practitioner perspective. Similarly, the consideration of other popular forecasting methods is an important issue as well.
- The interface between (and the potential of combining) temporal and cross-sectional aggregation has received minimal attention both in academia and industry and this is an issue that we plan to investigate in the next steps of our research.
- The extension of the work described here to cover inventory/implication metrics would allow a linkage between forecasting and stock control.
- Finally, the analytical and empirical consideration of Integer ARMA (INARMA) processes offers a great opportunity for advancements in the area of aggregation. Such processes bear a considerable relevance to intermittent demands where the benefits of aggregation may be even higher due to the reduction of zero observations.

Acknowledgements

We would like to thank the AE and anonymous referees for their comments that greatly helped to improve our paper.

APPENDIX A: COVARIANCE BETWEEN THE NON-AGGREGATED DEMAND AND AGGREGATED FORECAST FOR THE MA(1) AND AR(1) PROCESSES

The covariance between the non-aggregated demand and the forecast of aggregated demand can be calculated as follows:

$$\begin{aligned} Cov(d_t, F_T) &= Cov(d_t, \sum_{k=1}^{\infty} \beta(1-\beta)^{k-1} D_{T-k}) = \beta Cov(d_t, D_{T-1}) + \\ &\quad \beta(1-\beta)Cov(d_t, D_{T-2}) + \beta(1-\beta)^2 Cov(d_t, D_{T-3}) + \dots \end{aligned} \quad (A-1)$$

For an MA(1) process, by substituting (3) into (A-1) and considering $Cov(d_t, d_{t-k}) = 0$ for all $k > 1$ we have:

$$Cov(d_t, F_T) = \beta Cov(d_t, d_{t-1}) = \beta \gamma_1 \quad (A-2)$$

For an AR(1) process, by substituting (12) into (A-2) and performing some simplifications, we have

$$Cov(d_t, F_T) = \frac{\beta \phi \gamma_0}{1 - \phi^m + \beta \phi^m} \times \frac{1 - \phi^m}{1 - \phi} \quad (A-3)$$

APPENDIX B: VARIANCE OF THE AGGREGATED FORECAST FOR THE MA(1) AND AR(1) PROCESSES

The variance of the aggregated forecast can also be determined as in (10) but with different parameters. In order to obtain the value of the variance of the forecast error, we need to calculate the covariance between the aggregated demand and its forecast, so we begin by deriving the covariance between the aggregated forecast and the demand in period T :

$$\begin{aligned} Cov(D_T, F_T) &= Cov(D_T, \sum_{k=1}^{\infty} \beta(1-\beta)^{k-1} D_{T-k}) = \beta Cov(D_T, \sum_{k=1}^{\infty} (1-\beta)^{k-1} D_{T-k}) = \\ &\beta (Cov(D_T, D_{T-1}) + (1-\beta)Cov(D_T, D_{T-2}) + (1-\beta)^2 Cov(D_T, D_{T-2}) + \dots) \end{aligned} \quad (B-1)$$

The variance of the forecast after aggregation can be derived as:

$$\begin{aligned} Var(F_T) &= Var(\beta D_{T-1} + (1-\beta)F_{T-1}) = \beta^2 Var(D_{T-1}) + (1-\beta)^2 Var(F_{T-1}) \\ &\quad + 2\beta(1-\beta)Cov(D_{T-1}, F_{T-1}) \end{aligned} \quad (B-2)$$

For the MA(1) process, by substituting $Cov(D_T, D_{T-1}) = \gamma'_1$ and $Cov(D_T, D_{T-k}) = 0$ for all $k > 1$ into (B-1), we get:

$$Cov(D_T, F_T) = \beta\gamma'_1 \quad (B-3)$$

For the AR(1) process, by substituting $Cov(D_T, D_{T-1}) = \gamma'_1$ and $Cov(D_T, D_{T-k}) = \phi'^k \gamma'_1$ for all $k > 1$ into (B-1) we get:

$$Cov(D_T, F_T) = \frac{\beta\gamma'_1}{1 - \phi' + \beta\phi'} \quad (B-4)$$

Then, by using the fact that $Var(F_T) = Var(F_{T-k})$, $Cov(D_T, F_T) = Cov(D_{T-k}, F_{T-k})$ for all $k \geq 1$ and the fact that $Cov(D_{T-k}, D_{T-k}) = Var(D_{T-k}) = \gamma'_0$ for all k (the properties of stationary process) we have:

a) For the MA(1) process by substituting (B-3) into (B-2):

$$Var(F_T) = \frac{\beta\gamma'_0 + 2\beta(1-\beta)\gamma'_1}{2 - \beta} \quad (B-5)$$

b) For the AR(1) process by substituting (B-4) into (B-2):

$$Var(F_T) = \frac{\beta\gamma'_0}{2 - \beta} + \frac{2\beta(1-\beta)\gamma'_1}{(2 - \beta)(1 - \phi' + \beta\phi')} \quad (B-6)$$

APPENDIX C. PROOF OF THEOREM 1

By considering $MSE_{BA}/MSE_{AA} > 1$ and some simplifications, the quadratic function given by (C-1) should be negative

$$(2\theta - 2m\theta)\beta^2 + (m^2\eta + m(1 + \theta^2) + 2m\theta)\beta - 2m^2\eta \quad (C-1)$$

where

$$\eta = \frac{\alpha + \alpha\theta^2 + 2\alpha\theta}{2 - \alpha}. \quad (C-2)$$

Moreover, by investigating the sign of (C-1) we can obtain the conditions under which MSE_{BA}/MSE_{AA} is smaller, equal and greater than one. Now, we verify if the quadratic function (C-1) has real roots. To do so, we define the discriminant Δ of (C-1) as follows

$$\Delta = (m^2\eta + m(1 + \theta^2) + 2m\theta)^2 + 8(2\theta - 2m\theta)m^2\eta, \quad (C-3)$$

Now we use the fact that $-1 < \theta < 1$, $0 < \alpha < 1$ and $m \geq 2$ to obtain the values of Δ . If $\Delta < 0$ it means that (C-1) has no real roots and if $\Delta > 0$ it means (C-1) has two real roots. We can show that Δ in (C-3) is always positive, therefore (C-1) has two different roots denoted by β_1 and β_2 , where β_1 is defined in (31) and

$$\beta_2 = \frac{\left(-(m^2\eta + m(1 + \theta^2) + 2m\theta) - \sqrt{\left(\frac{(m^2\eta + m(1 + \theta^2))^2}{+ 2m\theta} \right)^2 + 8(2\theta - 2m\theta)m^2\eta} \right)}{2(2\theta - 2m\theta)}. \quad (C-4)$$

We can show that if $\theta < 0$, β_2 is always smaller than zero and $0 < \beta_1 < 1$ or $1 < \beta_1$ and $\theta > 0$, β_2 is greater than one and $0 < \beta_1 < 1$ or $1 < \beta_1$.

We know that the sign of the (D-1) between the two roots β_1 and β_2 is opposite to the sign of A , where $A = (2\theta - 2m\theta)$ is the sign of the coefficient of β^2 , Otherwise it is that the same as the sign of A . Now by considering β_1 , β_2 and A that is positive for $\theta < 0$ and negative for $\theta > 0$, we determine the sign of (C-1). So we have

- If $\theta < 0$, β_2 is always smaller than zero. If $0 < \beta_1$ then (C-1) is negative in the interval $[\beta_2, \beta_1]$ and it is positive outside this interval.
- If $\theta > 0$, β_2 is greater than one and we can show that $0 < \beta_1 < \beta_2$ thus (C-1) is positive in the interval $[\beta_1, \beta_2]$ and it is negative outside this interval.

From the above expressions we can see that when $\beta < \beta_1$, (C-1) is negative, otherwise when $\beta > \beta_1$, it is positive and when $\beta = \beta_1$, (C-1) is equal to zero. Equivalently

- If $\beta < \beta_1$, the ratio of MSE_{BA}/MSE_{AA} is greater than one and consequently the aggregation approach outperforms non-aggregation approach.
- If $\beta = \beta_1$, the ratio of MSE_{BA}/MSE_{AA} is equal to one and both strategies perform equally.
- If $\beta > \beta_1$, the ratio of MSE_{BA}/MSE_{AA} is smaller than one and the non-aggregation approach outperforms the aggregation approach.

APPENDIX D: SELECTION PROCEDURE FOR THE AR(1) PROCESS

Considering $MSE_{BA}/MSE_{AA} > 1$ is equivalent to having the quadratic function (D-1) negative, which subsequently is equivalent to

$$\begin{aligned} & (\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)\beta^2 \\ & + ((1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta)\beta \\ & - 2m^2(1-\phi^m)(1-\phi) \end{aligned} \quad (D-1)$$

For the quadratic function given by (D-1), the value of the discriminant Δ and the roots β_1 and β_2 can be defined as follows:

$$\Delta = ((1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta)^2 + 8(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)(m^2(1-\phi^m)(1-\phi)\eta), \quad (D-2)$$

$$\beta_1 = \frac{(-((1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta) + \sqrt{\Delta})}{2(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)}, \quad (D-3)$$

$$\beta_2 = \frac{(-((1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta) - \sqrt{\Delta})}{2(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)}, \quad (D-4)$$

where

$$\xi_1 = \left(m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right), \quad (\text{D-5})$$

$$\xi_2 = \left(\sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right), \quad (\text{D-6})$$

$$\eta = \frac{(-3\alpha\phi + \alpha + \alpha^2\phi)}{(2-\alpha)(1-\phi + \alpha\phi)}. \quad (\text{D-7})$$

We define the coefficient of β^2 in (D-1) as follows

$$A = \phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta. \quad (\text{D-8})$$

■

1. if $\Delta < 0$ then the non-aggregation approach is always provides more accurate forecasts, otherwise
 - If $\beta_2 < \beta < \beta_1$ then the aggregation approach works better.
 - If $\beta = \beta_1 = \beta_2$ then both approaches are identical.
 - If $\beta > \beta_1$ and/or $\beta < \beta_2$ then the non-aggregation approach works better.

APPENDIX E: PROOF OF THEOREM 2 AND THEOREM 3

Case 1. Using the fact that $1/3 < \phi < 1$, $m \geq 2$ and by considering the optimal smoothing constant, $\alpha^* = (3\phi - 1)/2\phi$ used to calculate MSE_{BA} , we can show that the discriminant Δ defined in (D-2) is negative, so there is no real root for (D-1). Consequently, the sign of (D-1) is the same as the sign of A defined in (D-2), we can show that the sign of A is always positive, therefore (D-1) is always positive and MSE_{BA}/MSE_{AA} is smaller than one. Hence, the non-aggregation approach always works better for the whole range of β and for any value of the aggregation level, m .

Case 2. $-1 < \phi \leq 1/3$. Using the fact that $-1 < \phi \leq 1/3$, $m \geq 2$ and by considering the small value of the smoothing constant before aggregation, $\alpha^* < 0.05$, it is straightforward to show that the discriminant Δ defined in (D-2) is positive, so (D-1) has two different roots denoted by β_1 and β_2 defined in (D-3) and (D-4) respectively.

We can show that the value of β_2 is either less than zero or greater than one. Now by considering the roots β_1 , β_2 and the sign of A , where A is defined in (D-8), we can determine the sign of (D-1) and consequently show the superiority of each approach.

- If $\beta_2 < 0$ and $\beta_1 > 0$, then (D-1) is negative in the interval $[\beta_2, \beta_1]$ and it is positive outside this interval.
- If $\beta_2 > 1$, we can show that $0 < \beta_1 < \beta_2$ and (D-1) is positive in the interval $[\beta_1, \beta_2]$ and it is negative outside this interval.
- Now from the above expressions we can get the following results:
 - If $\beta < \beta_1$, then $MSE_{BA}/MSE_{AA} > 1$.
 - If $\beta = \beta_1$, then $MSE_{BA}/MSE_{AA} = 1$.
 - Otherwise, $MSE_{BA}/MSE_{AA} < 1$.

REFERENCES

1. A. Chen and J. Blue, *Performance analysis of demand planning approaches for aggregating, forecasting and disaggregating interrelated demands*, International Journal of Production Economics **128** (2010), no. 2, 586-602.
2. A. A. Syntetos and J. E. Boylan, *The accuracy of intermittent demand estimates*, International Journal of Forecasting **21** (2005), no. 2, 303-314.
3. F. Chen, J. K. Ryan and D. Simchi-Levi, *The impact of exponential smoothing forecasts on the bullwhip effect*, Naval Research Logistics (NRL) **47** (2000), no. 4, 269-286.
4. U. Wemmerlov and D. C. Whybark, *Lot-sizing under uncertainty in a rolling schedule environment*, International Journal of Production Research **22** (1984), no. 3, 467-484.
5. A. Chen, C. H. Hsu and J. Blue, *Demand planning approaches to aggregating and forecasting interrelated demands for safety stock and backup capacity planning*, International Journal of Production Research **45** (2007), no. 10, 2269-2294.
6. K. Nikolopoulos, A. A. Syntetos, J. Boylan, F. Petropoulos and V. Assimakopoulos, *An aggregate-disaggregate intermittent demand approach (adida) to forecasting : An empirical proposition and analysis*, Journal of the Operational Research Society **62** (2011), no. 3, 544-554.
7. A. Silvestrini and D. Veredas, "Temporal aggregation of univariate and multivariate time series models: A survey," *Temi di discussione (working papers)*, Bank of Italy, Economic Research Department, 2008.
8. H. Widiarta, S. Viswanathan and R. Piplani, *Forecasting aggregate demand: An analytical evaluation of top-down versus bottom-up forecasting in a production planning framework*, International Journal of Production Economics **118** (2009), no. 1, 87-94.
9. M. Z. Babai, M. M. Ali and K. Nikolopoulos, *Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis*, OMEGA **40** (2012), no. 6, 713-721.
10. J. W. Taylor, *Exponential smoothing with a damped multiplicative trend*, International Journal of Forecasting **19** (2003), no. 4, 715-725.
11. E. S. Gardner, Jr., *Evaluating forecast performance in an inventory control system*, Management Science **36** (1990), no. 4, 490-499.
12. E. S. Gardner, Jr., *Exponential smoothing: The state of the art—part ii*, International Journal of Forecasting **22** (2006), no. 4, 637-666.
13. Y. Acar and E. S. Gardner Jr, *Forecasting method selection in a global supply chain*, International Journal of Forecasting **28** (2012), no. 4, 842-848.
14. T. Amemiya and R. Y. Wu, *The effect of aggregation on prediction in the autoregressive model*, Journal of the American Statistical Association **67** (1972), no. 339, 628-632.
15. G. C. Tiao, *Asymptotic behaviour of temporal aggregates of time series*, Biometrika **59** (1972), no. 3, 525-531.
16. G. Zotteri and M. Kalchschmidt, *A model for selecting the appropriate level of aggregation in forecasting processes*, International Journal of Production Economics **108** (2007), no. 1-2, 74-83.
17. H. Theil, *Linear aggregation of economic relations*, North-Holland Publishing Company, Amsterdam, Netherlands, 1954.
18. G. Yehuda and G. Zvi, *Is aggregation necessarily bad?*, Review of Economics and Statistics **42** (1960), no. 1, 1-13.
19. D. J. Aigner and S. M. Goldfeld, *Estimation and prediction from aggregate data when aggregates are measured more accurately than their components*, Econometrica **42** (1974), no. 1, 113-134.
20. S. Makridakis and M. Hibon, *The M3-competition: Results, conclusions and implications*, International Journal of Forecasting **16** (2000), no. 4, 451-476.
21. K. R. W. Brewer, *Some consequences of temporal aggregation and systematic sampling for arma and armax models*, Journal of Econometrics **1** (1973), no. 2, 133-154.
22. D. O. Stram and W. W. S. Wei, *Temporal aggregation in the arima process*, Journal of Time Series Analysis **7** (1986), no. 4, 279-292.
23. L. R. Souza and J. Smith, *Effects of temporal aggregation on estimates and forecasts of fractionally integrated processes: A monte-carlo study*, International Journal of Forecasting **20** (2004), no. 3, 487-502.
24. G. Athanasopoulos, R. J. Hyndman, H. Song and D. C. Wu, *The tourism forecasting competition*, International Journal of Forecasting **27** (2011), no. 3, 822-844.
25. E. S. Gardner, Jr. and E. McKenzie, *Forecasting trends in time series*, Management Science **31** (1985), no. 10, 1237-1237.
26. I. Luna and R. Ballini, *Top-down strategies based on adaptive fuzzy rule-based systems for daily time series forecasting*, International Journal of Forecasting **27** (2011), no. 3, 708-724.

27. K. Brannas, J. Hellstrom and J. Nordstrom, *A new approach to modelling and forecasting monthly guest nights in hotels*, International Journal of Forecasting **18** (2002), no. 1, 19-30.
28. M. Mohammadipour and J. E. Boylan, *Forecast horizon aggregation in integer autoregressive moving average (inarma) models*, OMEGA **40** (2012), no. 6, 703-712.
29. T. R. Willemain, C. N. Smart, J. H. Shockor and P. A. DeSautels, *Forecasting intermittent demand in manufacturing: A comparative evaluation of croston's method*, International Journal of Forecasting **10** (1994), no. 4, 529-538.
30. J. D. Croston, *Forecasting and stock control for intermittent demands*, Operational Research Quarterly (1970-1977) **23** (1972), no. 3, 289-303.
31. Spithourakis Georgios P., Fotios Petropoulos, Vassilios Assimakopoulos, Mohamed Zied Babai and K. Nikolopoulos, *Improving the performance of popular supply chain forecasting techniques* Supply Chain Forum: An International Journal **12** (2011), no. 4.
32. H. Widiarta, S. Viswanathan and R. Piplani, *On the effectiveness of top-down strategy for forecasting autoregressive demands*, Naval Research Logistics (NRL) **54** (2007), no. 2, 176-188.
33. H. Widiarta, S. Viswanathan and R. Piplani, *Forecasting item-level demands: An analytical evaluation of top-down versus bottom-up forecasting in a production-planning framework*, IMA Journal of Management Mathematics **19** (2008), no. 2, 207-218.
34. A. A. Syntetos., *"Forecasting of intermittent demand,"* vol. Phd, Business School, Buckinghamshire Chilterns University College, Brunel University, 2001.
35. W. W. S. Wei, *Time series analysis: Univariate and multivariate methods*, Addison-Wesley Publishing Company, Boston, USA, 2006.
36. N. Sariaslan, *"The effect of temporal aggregation on univariate time series analysis,"* Department of Statistics, Middle East Technical University, 2010, p. 150.
37. H. L. Lee, V. Padmanabhan and S. Whang, *Information distortion in a supply chain: The bullwhip effect*, Management Science **43** (1997), no. 4, 546-558.
38. H. L. Lee, K. C. So and C. S. Tang, *The value of information sharing in a two-level supply chain*, Management Science **46** (2000), no. 5, 626-643.
39. N. Erkip, W. H. Hausman and S. Nahmias, *Optimal centralized ordering policies in multi-echelon inventory systems with correlated demands*, Management Science **36** (1990), no. 3, 381-392.